

## Cyanogen Bromide (CNBr) Cleavage

Figure 1 shows the results of CNBr cleavage of a protein sample. The gel electrophoresis pattern on the left indicates the presence of a major band at approximately 66 Kd and a minor band at approximately 31 Kd.

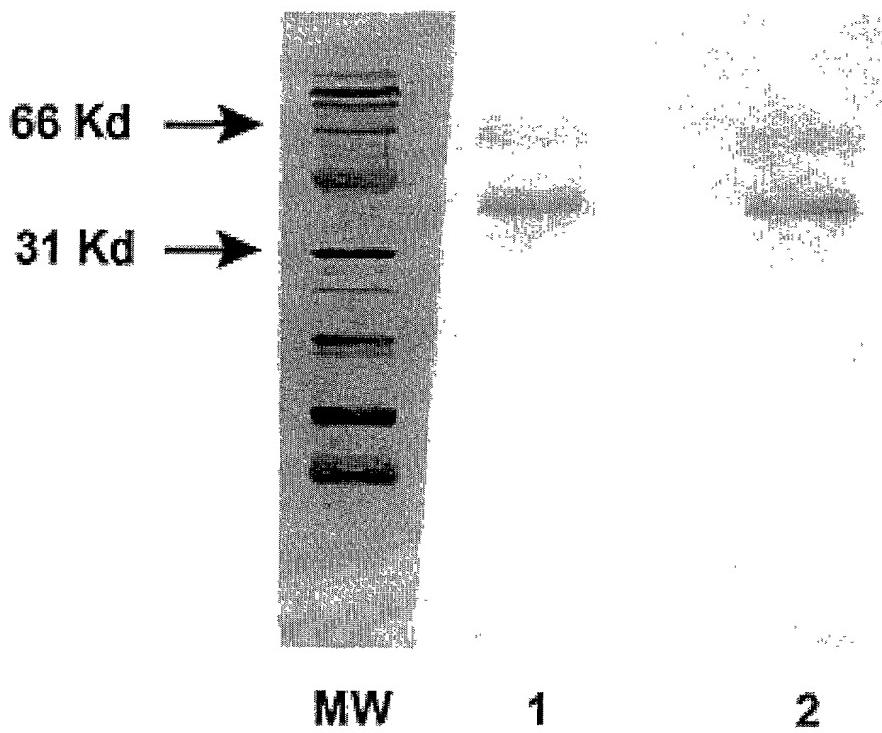
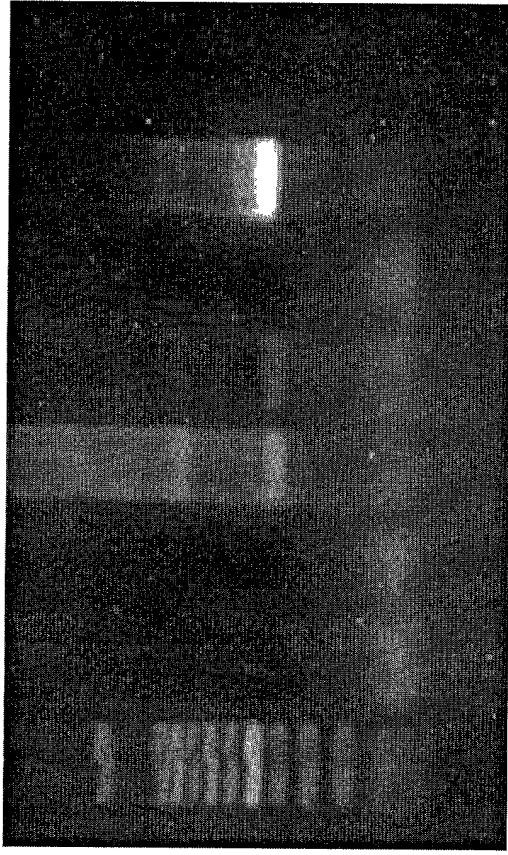


Figure 1

**A**

1 2 3 4 5 6



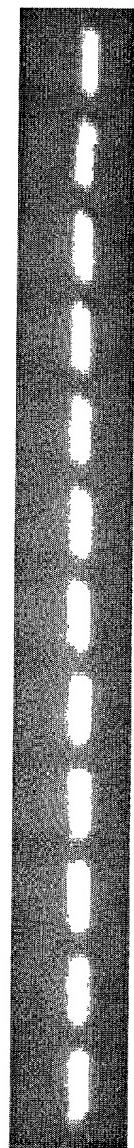
500 bases

**B**

1 2 3 4 5 6 7 8 9 10 11 12



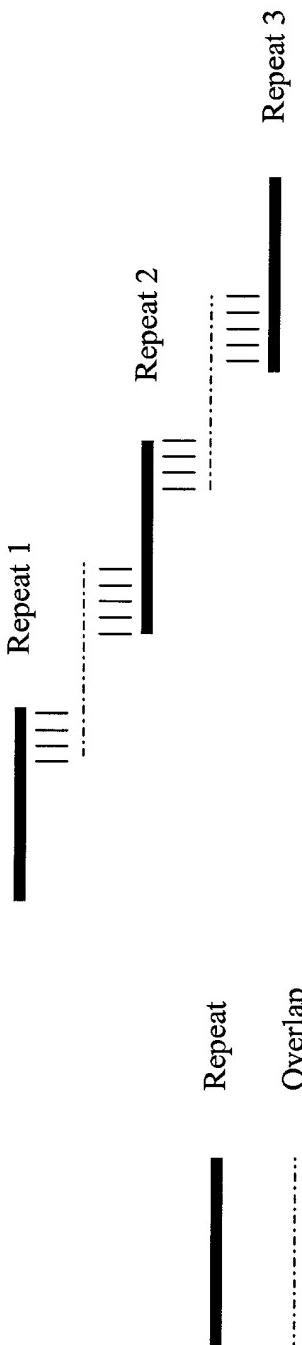
CA125



$\beta$ -Tubulin

**Figure 2**

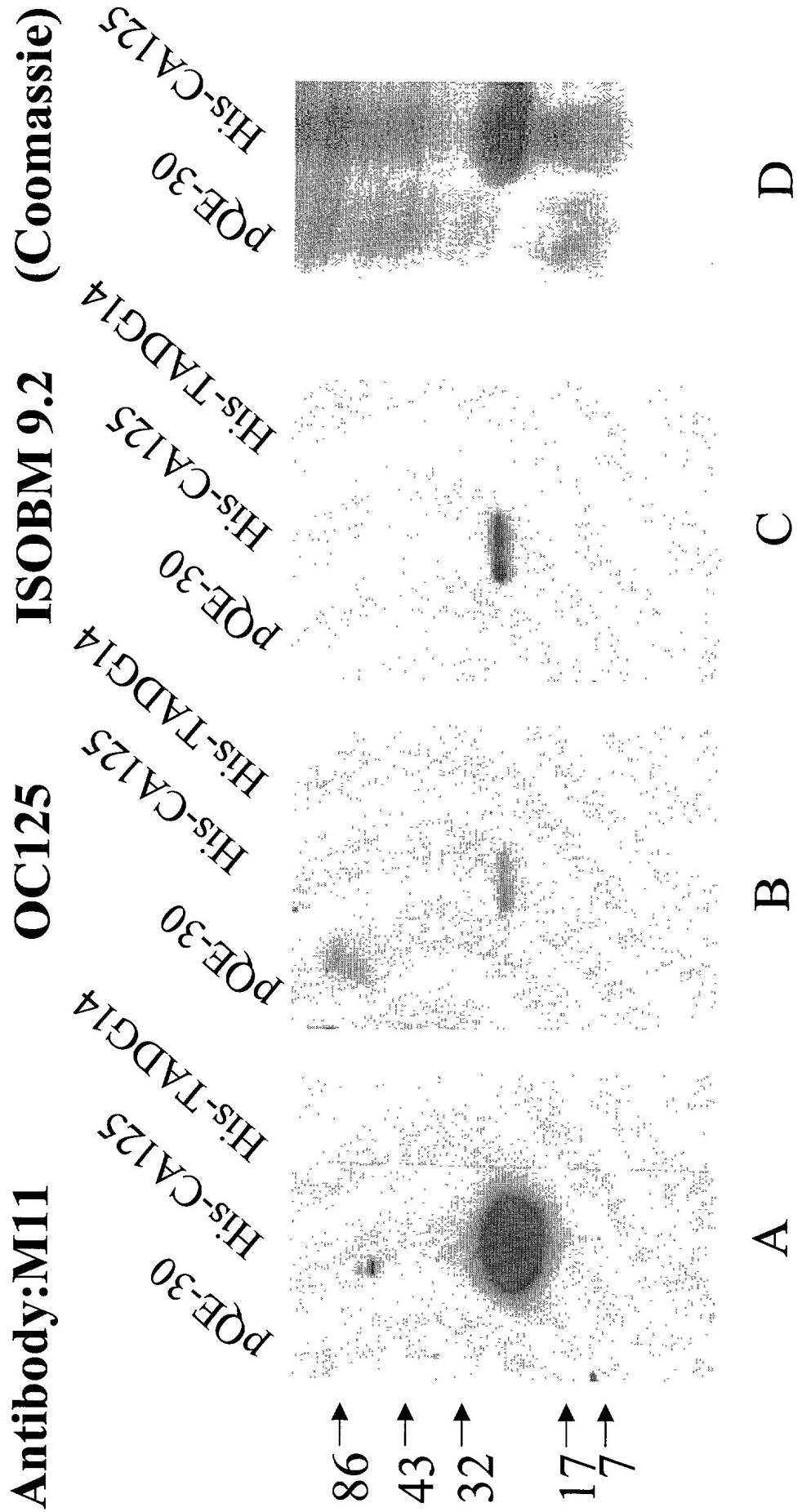
# A Strategy for Placing Repeat Sequences in Contiguous Order Using Overlap Sequence Alignment



(SEQ ID NO : 158)  
 1 AAVPMVYPTFLNETITNUQEEDMRHPPSRKFNATERELQGLKPLFRNSSLTLYSGCRLASLRPEKDSAMAVDAICTHRPDPEDLGLDRERLYWELNSLTINGIQELGPYTLDRNLSIYNGFTHSSMPTSTPGTSTVDGTGTCGTPSSSPST  
 156 2 AAGPLLVYPTFLNETITNUQEEDMRHPPSRKFNATERELQGLKPLFRNSSLTLYSGCRLASLRPEKDSAMAVDAICTHRPDPEDLGLDRERLYWELNSLTINGIQELGPYTLDRNLSIYNGFTHSSMPTSTPGTSTVDGTGTCGTPSSSPST  
 314 3 AAGPLLVYPTFLNETITNUQEEDMRHPPSRKFNATERELQGLKPLFRNSSLTLYSGCRLASLRPEKDSAMAVDAICTHRPDPEDLGLDRERLYWELNSLTINGIQELGPYTLDRNLSIYNGFTHSSMPTSTPGTSTVDGTGTCGTPSSSPST  
 470 4 AAGPLLVYPTFLNETITNUQEEDMRHPPSRKFNATERELQGLKPLFRNSSLTLYSGCRLASLRPEKDSAMAVDAICTHRPDPEDLGLDRERLYWELNSLTINGIQELGPYTLDRNLSIYNGFTHSSMPTSTPGTSTVDGTGTCGTPSSSPST  
 627 5 AAGPLLVYPTFLNETITNUQEEDMRHPPSRKFNATERELQGLKPLFRNSSLTLYSGCRLASLRPEKDSAMAVDAICTHRPDPEDLGLDRERLYWELNSLTINGIQELGPYTLDRNLSIYNGFTHSSMPTSTPGTSTVDGTGTCGTPSSSPST  
 781

(SEQ\_ID NO: 160)  
1 TAGPLLYPTFNTITNLYEEDMHPGSRRFNTTERVQGLLTPFKNTSVGFLYSGCRLTLLRPEKQEAATGVDTI CTHRVD PIGGLDRELYWELSOQLNSITELGPyTLDRLSLYNGFNPNSSVPTSTPGTSTVHIALTSGPSSPLGHT  
57 APVPLLPFTNFITDLYEENMHPGSRKNTTERVQGLLTPFKNTSVGFLYSGCRLTLLRPEKGAATGVDACTLRLDTPGGLDRELYWELSOQLNSTELGPyTLDRLSLYNGFTHSSSVPTSTPGTSAVHLETSGPSSPLGHT  
113 AFGPLLYPTFNTITNLYEEDMHPGSRKFTSTTERVQGLLTPFKNTSVSSLYSGCRLTLLRPERDGAATRVDACTRPDKSPGLDRELYWKLSQLTGHITELGPYTLDRHSLYNGFTHGSSMTTRPDTSTMHIALTSRPAISLSGPT  
169 TASPLLVLFNTITNQYEENMHPGSRKNTTERVQGLLTPFKNTSVGFLYSGCRLTLLRPKDGAATVDACTRPDKSPGLDRELYWELSOQLNSITELGPyTQDRDSLYNGFTHSSSVPTSTPGTSAVHLETSGPSSPLGHT  
156 312 468 624

Figure 3 (SEO ID NOS: 158, 159, 160, and 161)



**Figure 4**

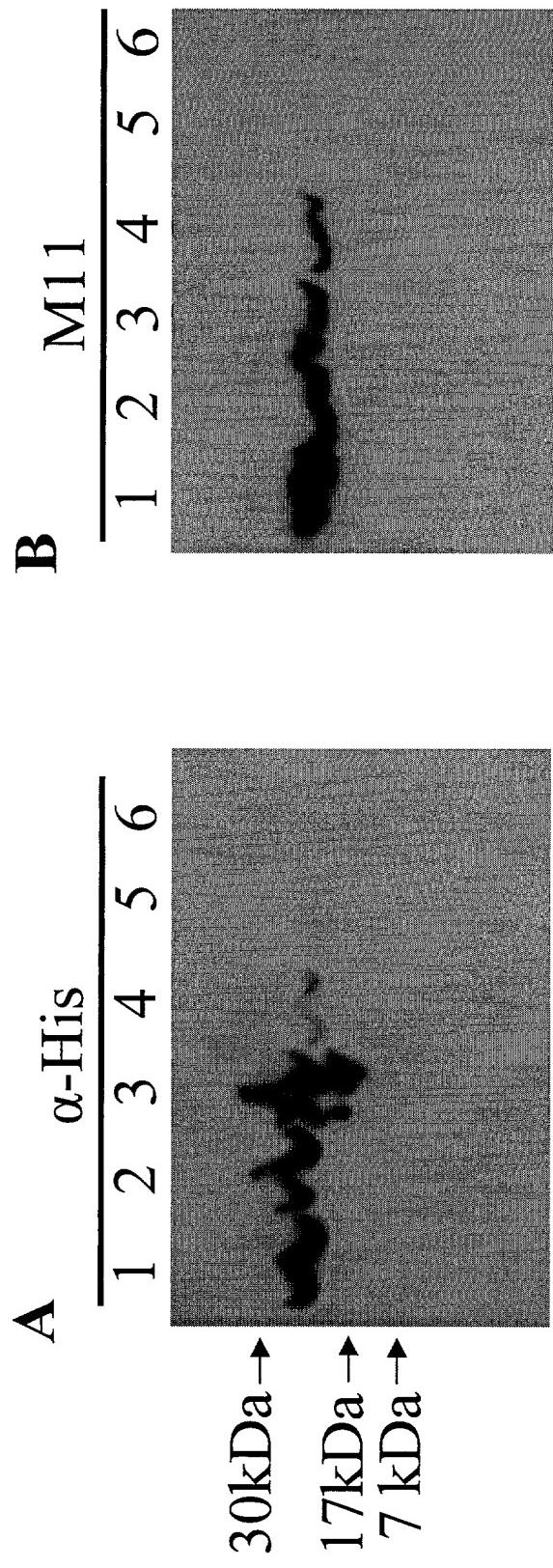


Figure 5 (SEQ ID NO: 150)

CA125 → N T 18S

CA125 → N T 18S

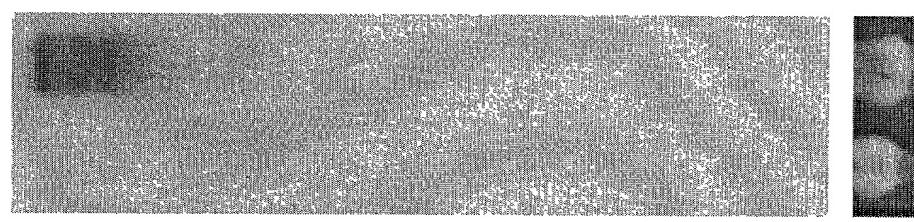


Figure 6

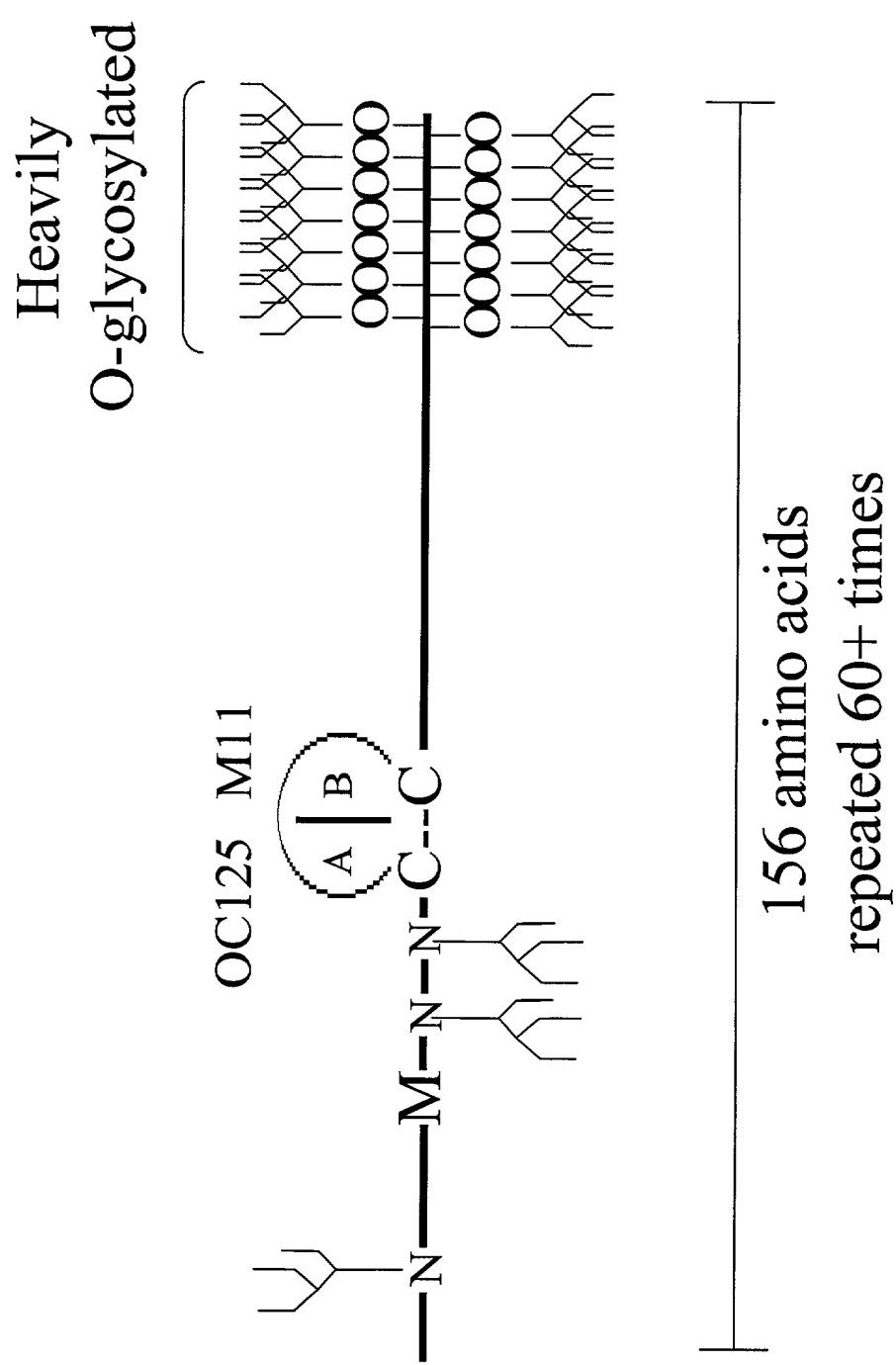


Figure 7A

W G D A C T E E G G G A V E D E E G G G A C G G

### Genomic Structure of a 156 Amino Acid Repeat Sequence of CA125

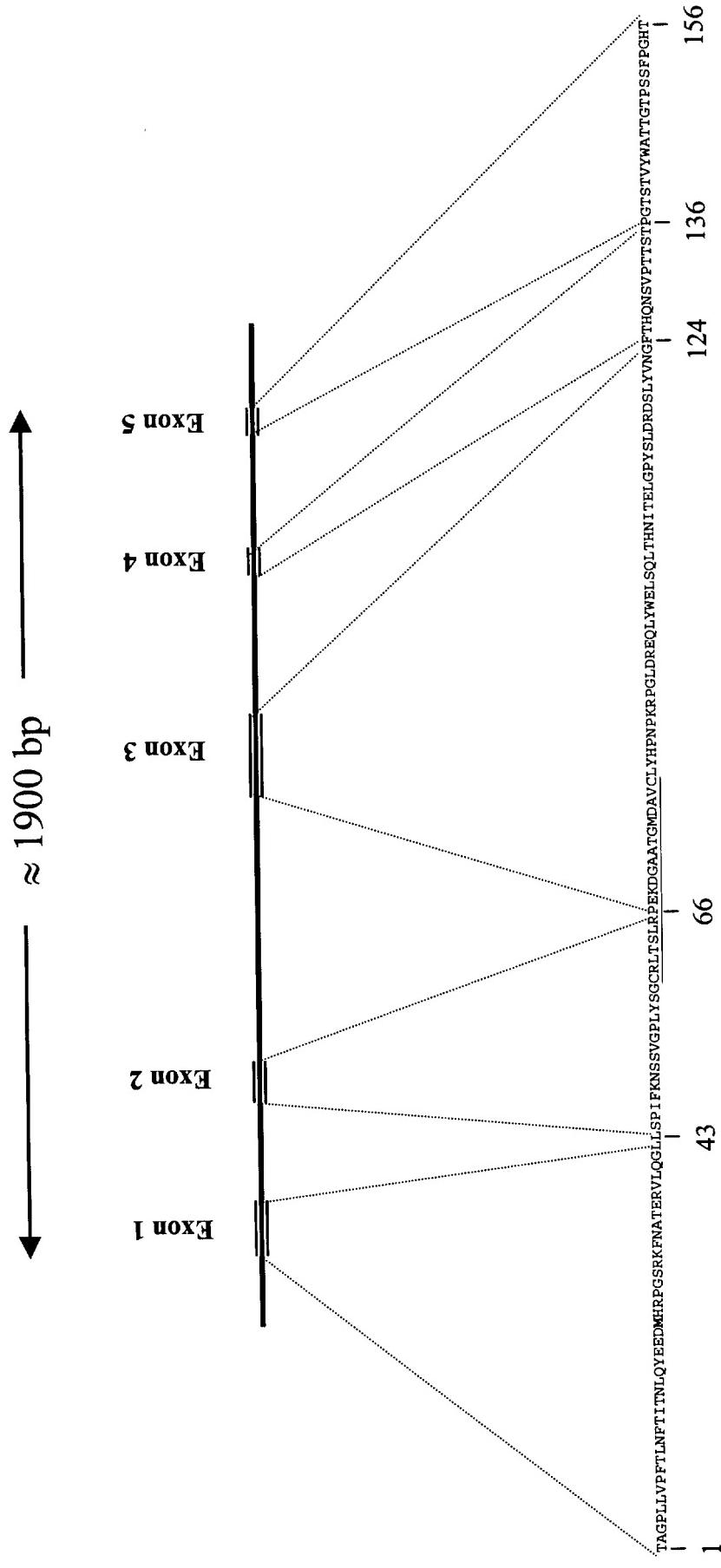


Figure 7B (SEQ ID NO: 163)

## Exon 1

1

42

ATVPFMVPFTLNFTITNLQYEDMRHPGSRKFNATERELQGL (SEQ ID NO: 164)  
TAVPLLPFTLNFTITNLQYGEDMRHPGSRKFNTTERVLQGL (SEQ ID NO: 165)  
VPGPLLVPFTLNFTITNLQYEEAMRHPGSRKFNTTERVLQGL (SEQ ID NO: 166)  
APGPLLVPFTLNFTITNLQYEDMRHPGSRKFNTTERVLQGL (SEQ ID NO: 167)  
APGPLLVPFTLNFTITNLQYEDMRHPGSRKFNTTERVLQGL (SEQ ID NO: 168)  
APGPLLVPFTLNFTITNLQYEDMRHPGSRKFNTTERVLQGL (SEQ ID NO: 169)  
SAGPLLVPFTLNFTITNLQYEDMRHPGSRKFNTTERVLQGL (SEQ ID NO: 170)  
AAGPLLMPFTLNFTITNLQYEDMRRTGSRKFNTMESVLQGL (SEQ ID NO: 171)  
TASPLLVLFINTCTITNLQYEDMRRTGSRKFNTMESVLQGL (SEQ ID NO: 172)  
AAGPLLVPFTLNFTITNLQYEDMRGHPGSRKFNTTERVLQGL (SEQ ID NO: 173)  
TAGPLLIPFTLNFTITNLQYGEDMGHPGSRKFNTTERVLQGL (SEQ ID NO: 174)  
TAGPLLVPFTLNFTITNLQYGEDMGHPGSRKFNTTERVLQGL (SEQ ID NO: 175)  
TAGPLLVLFTLNFTITNLKYEEDMHRPGSRKFNTTERVLQTL (SEQ ID NO: 176)  
TAGPLLVPFTLNFTITNLQYEDDMHRPGSRKFNATERVLQGL (SEQ ID NO: 177)  
TAGPLLVPFTLNFTITNLQYEDDMHRPGSRRFNTTERVLQGL (SEQ ID NO: 178)  
TAGPLLVPFTLNFTITNLQYEDDMHRPGSRKFNTTERVLQGL (SEQ ID NO: 179)  
APVPLLIPFTLNFTITNLQYEDDMHRPGSRKFNTTERVLQGL (SEQ ID NO: 180)  
ATGPVLLPFTLNFTITNLQYEDDMHRPGSRKFNTTERVLQGL (SEQ ID NO: 181)  
AAGPLLVPFTLNFTITNLQYEDDMHHPGSRKFNTTERVLQGL (SEQ ID NO: 182)  
SAGPLLVPFTLNFTITNLQYEDDMHHPGSRKFNTTERVLQGL (SEQ ID NO: 183)  
TASPLLVLFINTCTITNQRYEEENMHPGSRKFNTTERVLQGL (SEQ ID NO: 184)  
TASPLLVLFINTCTITNLRYEEENMHPGSRKFNTTERVLQGL (SEQ ID NO: 185)  
EPGPLLIPFTFNFTITNLHYEEENMQHPGSRKFNTTERVLQGL (SEQ ID NO: 186)  
EPGPLLIPFTFNFTITNLRYEEENMQHPGSRKFNTTERVLQGL (SEQ ID NO: 187)  
APVPLLIPFTLNFTITNLHYEEENMQHPGSRKFNTTERVLQGL (SEQ ID NO: 188)  
APVPLLIPFTLNFTITDLHYEEENMQHPGSRKFNTTERVLQGL (SEQ ID NO: 189)  
AASPLLVLFNLNGTITNLRYEEENMQHPGSRKFNTTERVLQGL (SEQ ID NO: 190)  
TAGPLLVPFTLNFTITNLKYEEDMHCPGSRKFNTTERVLQSL (SEQ ID NO: 191)  
AASHLLILFTLNFTITNLRYEEENMW.PGSRKFNTTERVLQGL (SEQ ID NO: 192)  
TGVVSEEPFTLNFTINNLRYMADMGMQPGSLKFNITDNVMKHL (SEQ ID NO: 193)  
AMGYHLKTLTLNFTISNLQYSQPMGKGSATFNSTEGVLQHLL (SEQ ID NO: 194)

Figure 7C

## **Exon 2**

43	65
LKPLFRNSSLSEYLYSGCRLASLR	(SEQ ID NO: 195)
LKPLFKNTSVSSLYSGCRLTLLR	(SEQ ID NO: 196)
LKPLFKNTSVGPLYSGCRLTLLR	(SEQ ID NO: 197)
LKPLFKSTSVGPLYSGCRLTLLR	(SEQ ID NO: 198)
LKPLFKNTSVGPLYSSCRLTLLR	(SEQ ID NO: 199)
LKPLFKNTSVGPLYSGCRLTSLR	(SEQ ID NO: 200)
LGPIFKNTSVGPLYSGCRLTSLR	(SEQ ID NO: 201)
LGPMFKNTSVGLLYSGCRLTLLR	(SEQ ID NO: 202)
LGPMFKNTSVGPLYSGCRLTLLR	(SEQ ID NO: 203)
LGPMFKNTSVGPLYSGCRLTSLR	(SEQ ID NO: 204)
LGPLFKNSSVGPLYSGCRLISLR	(SEQ ID NO: 205)
LGPLFKNSSVDPLYSGCRLTSLR	(SEQ ID NO: 206)
LSPIFKNSSVGPLYSGCRLTSLR	(SEQ ID NO: 207)
LSPIFKNTSVGPLYSGCRLTLLR	(SEQ ID NO: 208)
LSPLFQRSSLGARYTGCRVIALR	(SEQ ID NO: 209)
LRPLFKNTSVSSLYSGCRLTLLR	(SEQ ID NO: 210)
LRPLFKNTSVGPLYSGSRLTLLR	(SEQ ID NO: 211)
LRPLFKNTSIGPLYSSCRLTLLR	(SEQ ID NO: 212)
LRPLFKSTSVGPLYSGCRLTLLR	(SEQ ID NO: 213)
LRPVFKNTSVGLLYSGCRLTLLR	(SEQ ID NO: 214)
LRPVFKNTSVGPLYSGCRLTLLR	(SEQ ID NO: 215)
LRSLFKSTSVGPLYSGCRLTLLR	(SEQ ID NO: 216)
LRSLFKSTSVGPLYSGCRLTSLR	(SEQ ID NO: 217)
LTPLFKNTSVGPLYSGCRLTLLR	(SEQ ID NO: 218)
LTPLFRNTSVSSLYSGCRLTLLR	(SEQ ID NO: 219)
LMPLFKNTSVSSLYSGCRLTLLR	(SEQ ID NO: 220)
RPLFQKSSM.GPFYLGQLISLR	(SEQ ID NO: 221)

**Figure 7C**

### Exon 3

66

123

PEKDSSAMAVDAICTHRPDPEDLGLDRERLYWELSNLNTNGIQELGPYTLDRNSLYVNG (SEQ ID NO: 222)  
PEKDGAATGVDAICTHRLDPKSPGLNRQQLYWELSKLTNDIEELGPYTLDRNSLYVNG (SEQ ID NO: 223)  
PKKDGAATGVDAICTHRLDPKSPGLNRQQLYWELSKLTNDIEELGPYTLDRNSLYVNG (SEQ ID NO: 224)  
PEKDGATGVDAICTHHPDPKSPRILDREQQLYWELSQLTHTNITELGHYALDNDLSLVNG (SEQ ID NO: 225)  
PEKDGEATGVDAICTHRPDPGPGLDRREQQLYLELSQLTHTSITELGPYTLDRDSLYVNG (SEQ ID NO: 226)  
PEKDGAATGMDAVCLYHPNPKRPGLDREQQLYWELSQLTHTNITELGPySLDRDSLYVNG (SEQ ID NO: 227)  
PEKDGAATGMDAVCLYHPNPKRPGLDREQQLYCELSQLTHTNITELGPySLDRDSLYVNG (SEQ ID NO: 228)  
PEKDGAATRVDAACTYRPDPKSPGLDRREQQLYWELSQLTHTSITELGPYTLDRVSLYVNG (SEQ ID NO: 229)  
PKKDGAATKVDAICTYRPDPKSPGLDRREQQLYWELSQLTHTSITELGPYTQDRDSLYVNG (SEQ ID NO: 230)  
PKKDGAATKVDAICTHRLDPKSPGLDRREQQLYWELSQLTHTSITELGPYTLDRDSLYVNG (SEQ ID NO: 231)  
PEKDGAATRVDACTHRPDKSPGLDRERLYWKLSQLTHTGITELGPyTLDRHSLYVNG (SEQ ID NO: 232)  
PEKGVATRVDAICTHRPDPKIPGLDRQQQLYWELSQLTHTSITELGPYTLDRDSLYVNG (SEQ ID NO: 233)  
SEKDGAATGVDAICIHHDLPKSPGLNRERLYWELSQLTNGIKELGPYTLDRNSLYVNG (SEQ ID NO: 234)  
SEKDGAATGVDAICTHRLDPKSPGLDRREQQLYWELSQLTNGIKELGPYTLDRNSLYVNG (SEQ ID NO: 235)  
SEKDGAATGVDAICTHRLDPKSPGVDRREQQLYWELSQLTNGIKELGPYTLDRNSLYVNG (SEQ ID NO: 236)  
SEKDGAATGVDAICTHRVDPKSPGVDRREQQLYWELSQLTNGIKELGPYTLDRNSLYVNG (SEQ ID NO: 237)  
SEKDGAATGVDAICTHHLNPQSPGLDRREQQLYWQLSQMTNGIKELGPYTLDRNSLYVNG (SEQ ID NO: 238)  
PEKRGAAATGVDTICTHRLDPLNPGLDRREQQLYWELSKLTRGIIELGPYLLDRGSLYVNG (SEQ ID NO: 239)  
PEKNGAAATGMDAICSHRLDPKSPGLNRQQLYWELSQLTHTGIKELGPYTLDRNSLYVNG (SEQ ID NO: 240)  
PEKNGAAATGMDAICSHRLDPKSPGLDRREQQLYWELSQLTHTGIKELGPYTLDRNSLYVNG (SEQ ID NO: 241)  
PEKHGAATGVDAICTLRDPGPGLDRERLYWELSQLTNSVTELGPYTLDRDSLYVNG (SEQ ID NO: 242)  
PEKHGAATGVDAICTLRDPGPGLDRERLYWELSQLTNSITELGPYTLDRDSLYVNG (SEQ ID NO: 243)  
PEKHEAATGVDTICTHRVDPIGPGLDRERLYWELSQLTNSITELGPYTLDRDSLYVNG (SEQ ID NO: 244)  
PEKQEAATGVDTICTHRVDPIGPGLDRERLYWELSQLTNSITELGPYTLDRDSLYVNG (SEQ ID NO: 245)  
PEKQEAATGVDTICTHRVDPIGPGLDRERLYWELSQLTNSITELGPYTLDRDSLYVNG (SEQ ID NO: 246)  
PEKDKAATRVDAICTHHPDPQSPGLNRQQLYWELSQLTHTGITELGPyTLDRDSLYVNG (SEQ ID NO: 247)  
SVKNGAETRVDLLCTYLQPLSGPGLPIKQVFHELSQQTHGITRLGPYSLDKDSLNG (SEQ ID NO: 248)  
PEKDGAATGVDTTCYHPDPVGPGLDIQQLYWELSQLTGHVTQLGFYVLDRLDSLFING (SEQ ID NO: 249)

Figure 7C

<b>Exon 4</b>	<b>Exon 5</b>
124        135	136        156
FTHRSSMPTTST (SEQ ID NO: 250)	PGTSTVDVGTSGTPSSSPSPT (SEQ ID NO: 278)
FTHRSSMPTTSI (SEQ ID NO: 251)	PGTSTVDLRTSGTPSSLSSPTIM (SEQ ID NO: 279)
FTHRTSVPTSS (SEQ ID NO: 252)	PGTSTVDLGTSGTPFSLPSPA (SEQ ID NO: 280)
FTHRTSVPTTST (SEQ ID NO: 253)	PGTSTVDLG.SGTPSSLPSPT (SEQ ID NO: 281)
FTHRSSVPTTSS (SEQ ID NO: 254)	PGTSTVDLG.SGTPSLPSSPT (SEQ ID NO: 282)
FTHRSSVSTTST (SEQ ID NO: 255)	PGTSTVDLGTSGTPSSLPSPT (SEQ ID NO: 283)
FTHRSSVAPTTST (SEQ ID NO: 256)	PGTPTVDLGTSGTPVSKPGPS (SEQ ID NO: 284)
FTHRSSGLTTST (SEQ ID NO: 257)	PWTSTVDLGTSGTPSPVPSPPT (SEQ ID NO: 285)
FTHRSFGLTTST (SEQ ID NO: 258)	PGTSTVYWATTGTPSSFPGHT (SEQ ID NO: 286)
FTHRSSFLTTST (SEQ ID NO: 259)	PGTSTVHLATSGTPSSLPGHT (SEQ ID NO: 287)
FTHRNFVPITST (SEQ ID NO: 260)	PGTSTVHLATSGTPSPLPGHT (SEQ ID NO: 288)
FTHRSSVPTTSI (SEQ ID NO: 261)	PDTSTMHLATSRTPASLSGPT (SEQ ID NO: 289)
FHQSSVSTTST (SEQ ID NO: 262)	PGTSAVHLETSGTPASLPGHT (SEQ ID NO: 290)
FTHQTSAPNTST (SEQ ID NO: 263)	PGTSAVHLETTGTPSSFPGHT (SEQ ID NO: 291)
FTHQTFAPNTST (SEQ ID NO: 264)	PGTSTVHLGTSETPSSLPRPI (SEQ ID NO: 292)
FTHQNSVPTTST (SEQ ID NO: 265)	PGTSIVNLGTSGIPPSLPETT (SEQ ID NO: 293)
FTHQSSMTTRT (SEQ ID NO: 266)	PGTFTVQPETSETPSSLPGPT (SEQ ID NO: 294)
FTHWIPVPTTST (SEQ ID NO: 267)	PGTPTVDLGTSGTPVSKPGPS (SEQ ID NO: 295)
FTHWSPIPTTST (SEQ ID NO: 268)	PGTPTVYLGASKTPASIFGPS (SEQ ID NO: 296)
FTHWSSGLTTST (SEQ ID NO: 269)	PKPATTFLPPLEATT..... (SEQ ID NO: 297)
FHPRSSVPTTST (SEQ ID NO: 270)	QINFHIVNWNLNSNPDPTSSEY (SEQ ID NO: 298)
FNPRSSVPTTST (SEQ ID NO: 271)	
FNPWSSVPTTST (SEQ ID NO: 272)	
FTQRSSVPTTSI (SEQ ID NO: 273)	
FTQRSSVPTTST (SEQ ID NO: 274)	
FTQRSSVPTTSV (SEQ ID NO: 275)	
YNEPGLDEPPTT (SEQ ID NO: 276)	
YAPQNLISRGEY (SEQ ID NO: 277)	

**Figure 7C**

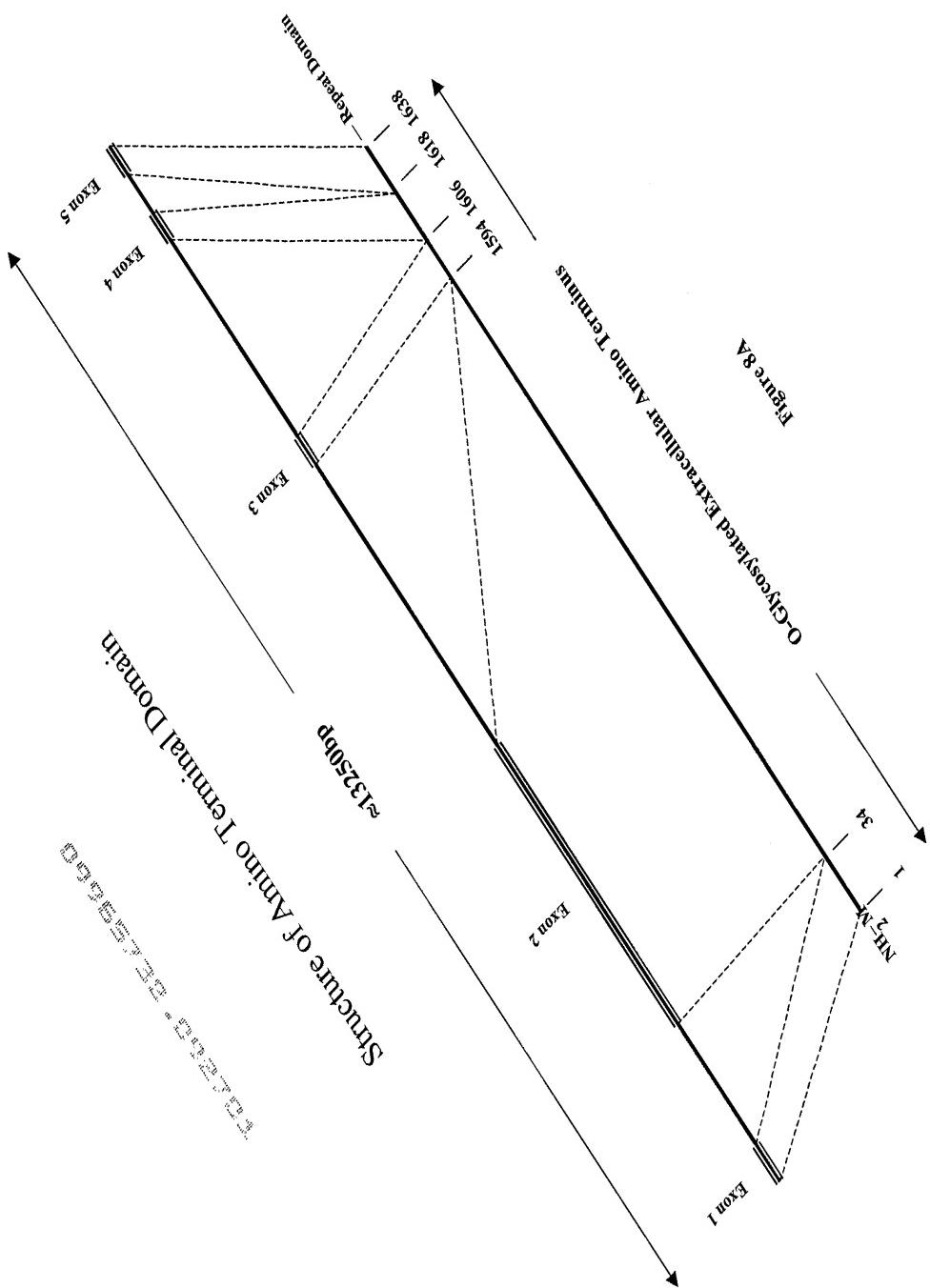


Figure 8B (SEQ ID NO: 299)  
TLDNRNSLYNN GETHRSSMPT TGTGPGTSTVD VGTSGTPSSS PSPPT

1	<b>MEHITKIPNE</b>	AHRG <small>T</small> IRPV	KGPQTSTSPA	SPKGLHTGGT	KRMETTTAL	901	ISATFPTVPE	SPESEATAS	WVTHPAVTST	TVPRPNTYS	HSEPDTTIPS <small>I</small>
51	<b>KTTTALKTT</b>	SRATLTTSVY	TPTLGLTTPL	<b>N</b> ASRQMASTI	LTEMMITTPY	951	OO	OO	DFPTITVSPD	VPMVTSQV <small>F</small>	SSGTDTSITI
101	<b>VFPDVPERTS</b>	SLATSLGAET	STALPRRTTPS	VLNRESETTA	SLVRSGAER	1001	OO	OO	LPVSPGASKM	ITSUVISSG <small>T</small>	DSTTFRP <small>T</small> LT
151	<b>SPVIQTLDS</b>	SSEPTTASW	VIHPAETIPT	VSKTTPNFFH	SELDTVSSTA	1051	ETTISFITYS	ETHISSAIP <small>T</small>	EPYEPETTA	TQLIHPAETN	SHSKSDTTL
201	<b>TSHGADVSSA</b>	IPTNISPSS	DALTPLVTIS	GTDTSTTFFP	LTKSPHETET	1101	OO	OO	PDMSDLVTS	VPSGGTDTST	TEPTLSETPY
251	RTTWLTHPAE	TSSTIPTTIP	NFSHESDAT	PSIATSPGAE	TSSALPIMTV	1151	HPAETSTTVS	GTIPNFSHRG	SDTAPSMTS	PGVDTRSQGV <small>P</small>	TTTIPPSIPG
301	<b>SPGAEDLVTS</b>	QVTSSGTDRN	MTIPTLTLSP	GEPTKIASLV	THPEAQOTSSA	1201	VVTSQVTSSA	TDTSIAITPL	TPSPGEPEPT	ASSATHPGTQ	TGFTVPIRT
351	IPSTTISPAV	SRLVTSMWTS	LAARTSTTNR	ALTNSPGEPA	TTVSLVTHPA	1251	PSSEPDTMAS	WVTHPPQST	PVSRTSSFG	HSSPDATPVM	ATSPRTEASS
401	QTSPTVPWTI	SIFFHSKSDT	TPSMTTSHGA	ESSSAVEPTP	VSTEVPGVVT	1301	AVLTTISPGA	PEMVTQSITS	SGAATSTTV	TLTHSPGMPE	TTALLSTHPR
451	PLVTSSRAVI	STTIPILTS	PGEPEPTPSM	ATSHGEAASS	AIPTPTVSPG	1351	O	O	TVEPVQVSETT	ASLTIRGAE	TSTALPTQTT
501	VPGVVTSLVT	SSRAVSTTTI	PILTFLSLGEP	ETTPSMASH	GTEAGSAVPT	1401	TSRVDSLSP <small>T</small> A	SPGVSAKTAP	LSTHPGTETS	TMIPSTL	SLLETTGLLA
551	VLPEVPGMVT	SLVASSRAVT	STTIPLTLS	PGEPEPTPSM	ATSHGAEASS	1451	O	O	TSSSAETSTS	TLTLTVSPAV	SGLSSASIT
601	TVPTVSP EVP	GVVTSLVTS	SGVNNTSIPT	LILSPGELET	TPSMATSHGA	1501	GPPEFSRTVT	GTTMLTIPSE	MPTPPKTSHG	EGVSPTTILR	TMVEATNL
651	EASSAVPTPT	VSPGVGSVVT	PLVTSRRAVT	STTIPLTLS	SSEPETTPSM	1551	OO	OO	SLFTPLTTPG	MSTLASESVT	SRTSYNHRSW
701	ATSHGVEASS	AVLTVSP EVP	GMVTSLVTS	RAVTSTTIP	LTISSEDEPET	1601	ITSTSYNRR	YWTPAT	SSIPSTAAT	VPMVPTLN	X
751	TTSLVTHSEA	KMISALPTLA	VSPTVQGLVT	SLVSSGSET	SAFSNLTVAS	1651	FTITNLQYE	DMRHPSRKF	NATERELQGL	LKPLFRNSSL	EYLYSGCRLA
801	SQFETIDSWV	AHPGTEASSV	VPTLTVSTGE	PFTNISLYTH	PAESSSTLLR	1701	SLRPEKDSSA	MAVDAILCTHR	PDPEDLGDR	ERYWELSNL	TNGIQELGPY
851	TTSRFSHSEL	DTMPSTVTS	EAESSAIST	TISPGIPGV <small>L</small>	TSLVTSGRD	1751					

Figure 8B (SEQ ID NO: 299)

## Structure of Carboxy Terminal Domain

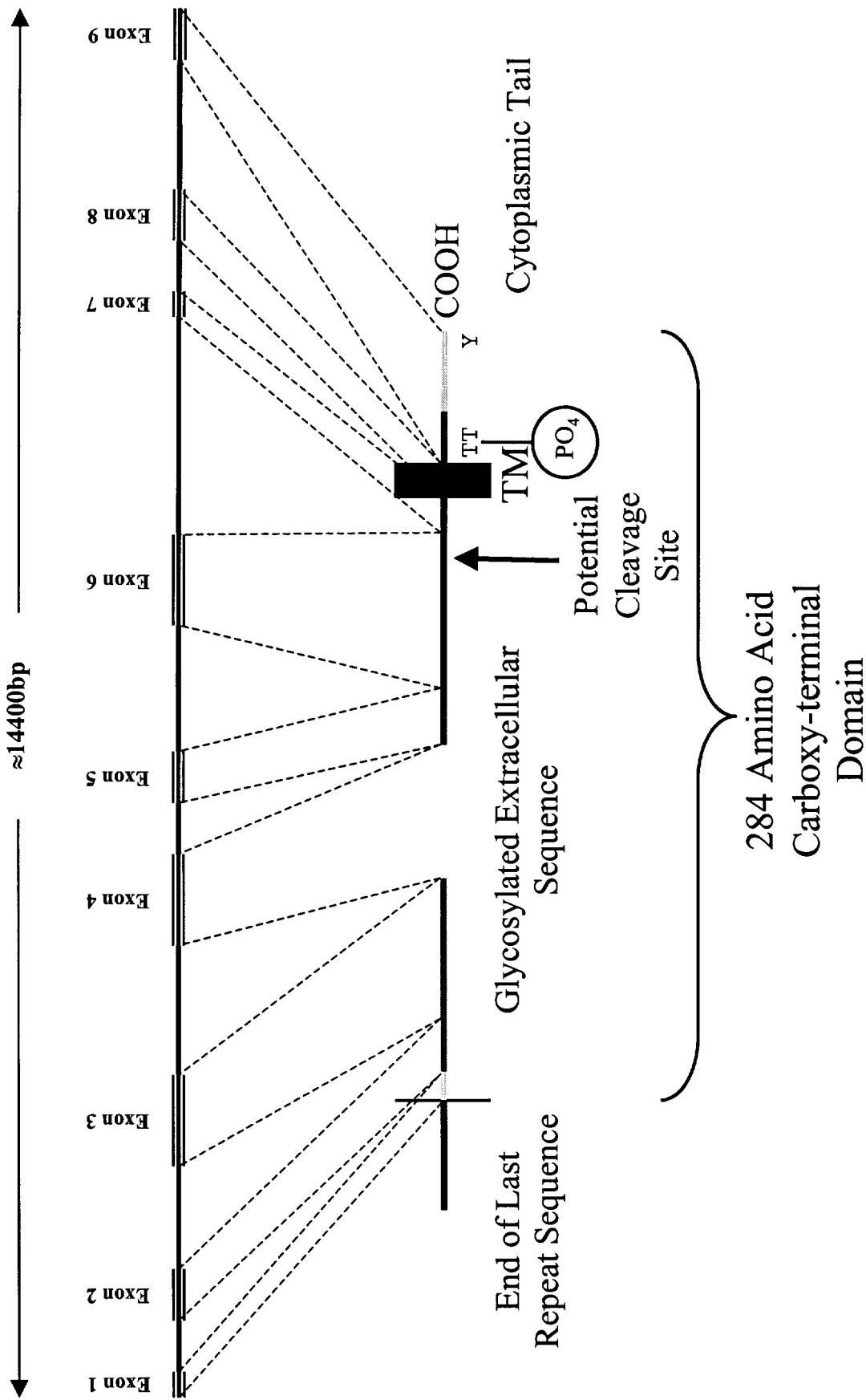


Figure 9A

ITLLRDIQDK VTTLYKGSQL HDTFRFCLVLT NLTMDSVLVT VKALFSSNLD

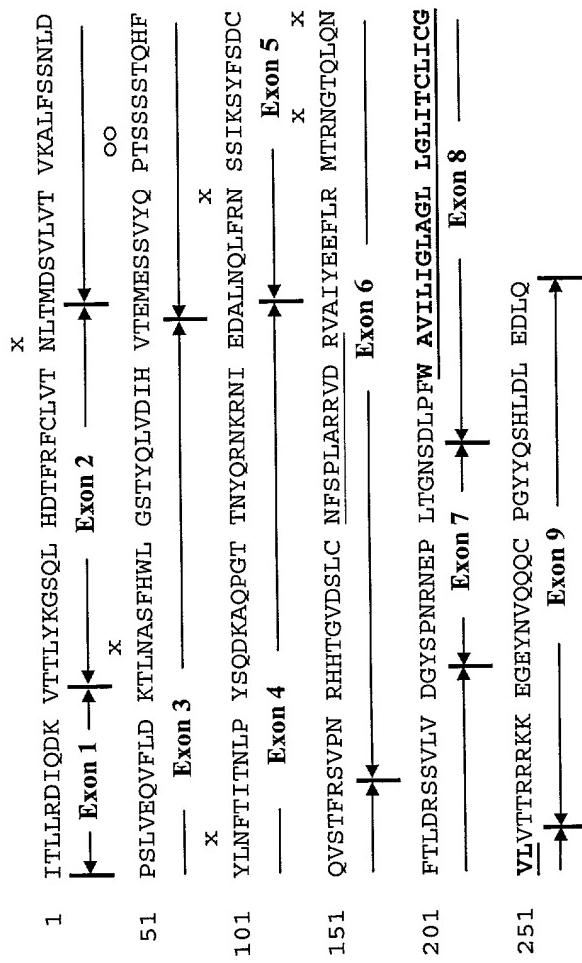


Figure 9B (SEQ ID NO: 300)

# Proposed Structure of CA125

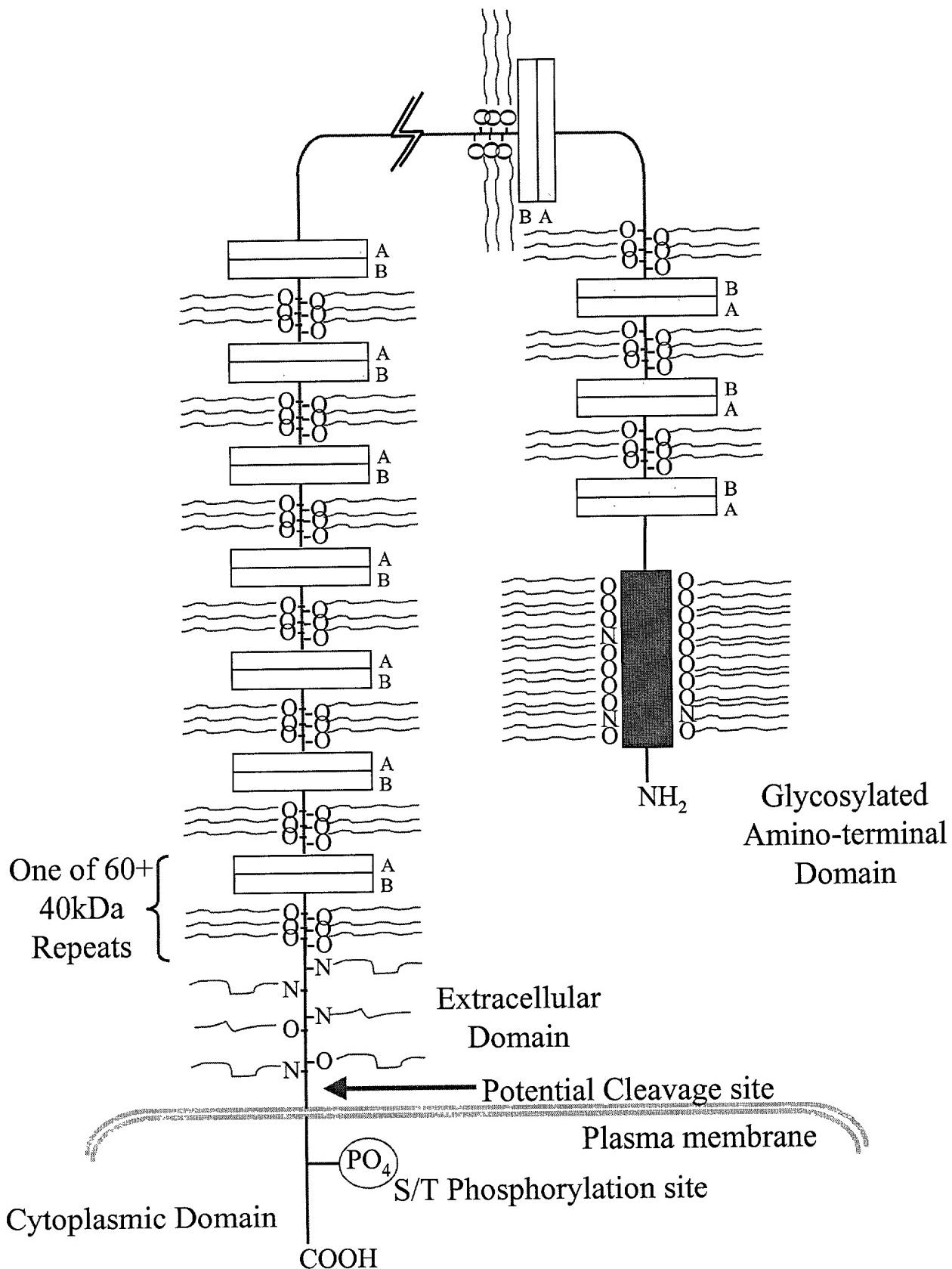


Figure 10